Modelling Interaction with Economic Models of Search

Leif Azzopardi
School of Computing Science, University of Glasgow
Glasgow, United Kingdom
Leif.Azzopardi@glasgow.ac.uk

ABSTRACT

Understanding how people interact when searching is central to the study of Interactive Information Retrieval (IIR). Most of the prior work has either been conceptual, observational or empirical. While this has led to numerous insights and findings regarding the interaction between users and systems, the theory has lagged behind. In this paper, we extend the recently proposed search economic theory to make the model more realistic. We then derive eight interaction based hypotheses regarding search behaviour. To validate the model, we explore whether the search behaviour of thirty-six participants from a lab based study is consistent with the theory. Our analysis shows that observed search behaviours are in line with predicted search behaviours and that it is possible to provide credible explanations for such behaviours. This work describes a concise and compact representation of search behaviour providing a strong theoretical basis for future IIR research.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval:Search Process; H.3.4 [**Information Storage and Retrieval**]: Systems and Software:Performance Evaluation

General Terms

Theory, Experimentation, Economics, Human Factors

Keywords

Retrieval Strategies, Search Behaviour, Evaluation

1. INTRODUCTION

How information seekers behave and interact with Information Retrieval systems is a fundamental question in the area of Interactive Information Retrieval [10]. Considerable empirical and observational research has been undertaken which has led to various findings about users, their behaviours, their perceptions of systems and system performance [21, 23, 33, 36, 37, 38, 43, 44]. For example with respect to querying behaviour, it has been observed that users worked harder on systems that were less effective by posing more queries [33, 38] and that they could achieve their goal by posing a series of short queries [23]. However, users issued fewer queries when the cost of querying increased [6, 13], while more experienced users posed fewer queries than inexperienced users [43]. Such observations are valuable, helping to piece together how users behave and act under various circumstances. However, it is difficult to link such findings together as much of the empirical and observational studies have been performed independently, and have not be guided by any underlying theory [22]. Consequently, developing formal models and theories that describe, predict and explain search behaviour has been hailed as one of the grand challenges of IIR [10, 18].

While numerous models of the information retrieval and seeking process have been developed (e.g. [8, 9, 11, 15, 17, 24, 46]), they have been largely conceptual in nature showing where researchers should focus their attention. However, such models do not make predictions or explain observed search behaviours, i.e. they are descriptive [20]. For example, a popular descriptive model is Bates' Berry Picking metaphor which described users as foragers, who go from patch to patch, choosing the best berries from the bushes in each patch [8]. While this metaphor might be an apt description of how people interact with search results, it does not provide any insight into why people behave like this, nor does it help predict how people will behave under different circumstances. In [20], such models are referred as pretheoretical, in the sense that they point out the relationships between factors which can be used to develop more formal and predictive models of interaction from which hypotheses can be generated. While these lines of research have been useful, the focus has mainly been on answering the question: how people behave when searching? However, a fundamental question persists: *why do people behave in such ways?*

A promising development in the late 1990s was the introduction of Information Foraging Theory (IFT) [25], which sought to predict and explain various information interactions. Under IFT, for example, the berry picker's actions are quantified mathematically in order to make predictions about how they will behave, where it is assumed that the berry picker will seek to maximise the rate at which they acquire relevant information. The theory implies that the berry picker will stay longer in a patch when it takes them longer to get there (i.e. if the cost of a query increases, then

the user will examine more results). While IFT received a lot of initial interest, most research was focused on browsing [26, 27] rather than applied specifically to ad-hoc topic retrieval. There has recently however been a renewed interest in developing formal models of interaction which are mathematical and computational in nature [2, 3, 7, 6, 16, 45], that specifically focus on topic retrieval. In these works the relationship between the user's interactions, the associated costs, and the benefit obtained from the system is formalised. As a consequence, the use of such models can lead to directly testable hypotheses about search behaviour; either directly from the theory or via simulations. However, such models have been criticised because they often make numerous assumptions about users [4].

In this paper, we will be focusing on the model of search based on Economic theory [2, 5]. While the initial theory provided some interesting insights and explanations about how users behave [2], it makes a number of critical modelling assumptions which detracts from its realism and applicability. Furthermore, empirical research has shown that the current theory failed to convincingly explain the observed search behaviours when tested [5]. In this work, we propose a new model of search, addressing some of the limitations of the previous attempt. This leads to a number of deeper insights regarding search behaviours and a number of novel contributions: (1) we provide a better explanation of observed search behaviours, (2) we show how the model leads to a number of specific hypotheses about search behaviour, (3) we explain how the search behaviour of users will change as cost and performance change, and (4) we provide a comprehensive empirical analysis contrasting actual search behaviour of thirty-six participants against the theory developed. Finally, we show that the predicted relationships between cost, performance and interaction hold and the observed search behaviour is consistent with the theory.

2. ECONOMICS IN IR

Economics provides an array of tools for modelling decision making, dealing with risk and handling uncertainty [39]. Early Information Retrieval (IR) research exploited such tools to examine IR systems in a number of ways ranging from purchasing decisions [1, 29] to ranking [16, 28, 41] to user behaviour [2, 12, 14]. Initial attempts focused on the trade-off between the cost of an IR system and its effectiveness. In [1, 29], Axelrod and Rotheberg compare different mechanised IR systems available during the late 1960s and early 1970s by performing a cost benefit analysis in order to decide which system to purchase. In [14], Cooper took a more user-oriented perspective. He modelled the trade-off between the amount of time a user should spend searching versus how much time the system should spend searching. In the same period, Robertson [28] examined the problem of ranking in terms of the costs and benefits of ranking one document above another. This led to the formulation of the Probability Ranking Principle (PRP) which essentially applies decision theory to the ranking problem [28]. More recently, Fuhr revised and extended the PRP to consider a series of interactions in the interactive Probability Ranking Principle (iPRP) [16]. This generalised model accounted for the different costs and benefits associated with particular choices when ranking documents.

In [40], Varian outlined three directions in which economics could be useful for search: (1) to obtain better estimates of the probability of relevance, (2) to apply Stigler's theory on Optimal Search Behaviour to IR [35], and (3) to examine the economic value of information using consumer theory, "where a consumer is making a choice to maximise expected utility or minimise expected cost" [40]. A number of different works have begun to examine these directions. For example, in [41], Wang and Zhu used Portfolio theory to obtain better estimates of relevance by accounting for the uncertainty associated with the probability estimates when ranking. While in [12], Birchler and Butler explain how Stigler's theory can be applied to search in order to predict when a user should stop examining results in a ranked list. However, they did not conduct any empirical study to verify whether the theory was consistent with users actual behaviour. In a variation on Varian's third suggestion, Azzopardi suggested that Production Theory could be used to model the search process instead [2]. This led to the development of Search Economic Theory (SET) which has been specifically developed to model ad-hoc topic retrieval. However, there are a number of limitations and problems with this approach. In the next section, we shall provide an overview of the model and review the criticisms of the approach, before proposing a new economic model of search.

3. ECONOMIC MODEL OF SEARCH

In [2], the search process was modelled using an analogy to Production Theory [39]. Production Theory, also known as the theory of firms, models the situation where a firm takes inputs (such as capital and labour) and converts them to output (such as products or services). Applied to search, a user with a search engine is considered the firm, where the inputs are the user interactions and the output is the gain received from the relevant documents found during the search process. The model proposed abstracts the search process down to two main interactions: (i) querying, and (ii) assessing, and consists of two functions that characterise gain and cost. The gain function proposed in [2] was:

$$\mathbf{g}(\mathbf{Q}, \mathbf{A}) = \mathbf{k}.\mathbf{Q}^{\boldsymbol{\alpha}}.\mathbf{A}^{(\mathbf{1}-\boldsymbol{\alpha})} \quad (1)$$

where it is assumed that the total amount of gain (i.e. session based Cumulative Gain [19]) is determined by the number of queries issued ($\mathbf{Q}$) and the number of documents assessed per query ($\mathbf{A}$), $\mathbf{k}$ expresses the quality of the technology and how well it is used, while $\boldsymbol{\alpha}$ denotes the relative efficiency of querying versus assessing. This function is often referred to as the Cobbs-Douglas production function in Economics [40]. In [2], it was shown that simulated usage data on several standard retrieval models closely fitted this function. In Figure 1, we have plotted the gain curve for BM25 ($\mathbf{k} = 5.394$, $\boldsymbol{\alpha} = 0.576$, $ncg = 0.6$ [2]). Each point on the curve represents the same amount of gain, but for different combinations of the inputs, $\mathbf{Q}$ and $\mathbf{A}$.

The following cost function was then used to ascribe a cost to each combination of inputs:

$$\mathbf{c}(\mathbf{Q}, \mathbf{A}) = \mathbf{c_q}.\mathbf{Q} + \mathbf{c_a}.\mathbf{A}.\mathbf{Q} \quad (2)$$

where $\mathbf{c_q}$ is the cost of issuing a query, and $\mathbf{c_a}$ is the cost of assessing a document[1]. To show how the cost relates to gain,

[1] In [2], $\mathbf{c_a}$ was unit normalised, i.e. set to one, and $\mathbf{c_q}$ was called $\boldsymbol{\beta}$ the relative cost of a query to an assessment, such that $\boldsymbol{\beta} = \mathbf{c_q}/\mathbf{c_a}$

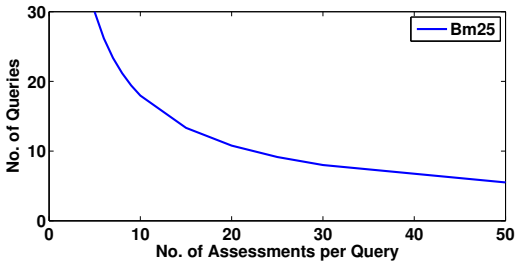

Figure 1: Plot of the Gain Function for BM25.

we have plotted the cost curve in Figure 2 where $\mathbf{c_q} = 4$ and $\mathbf{c_a} = 1$ for the gain curve above. According to the theory, users will seek to minimise their cost for a given level of gain (or alternatively maximise their gain for a given cost). By inspecting the cost curve plot, it is clear that the minimum cost is when $\mathbf{A} = 10$, which corresponds to when $\mathbf{Q} = 18$. Any other combination of inputs would result in a higher cost. In [5], the cost of a query was varied to show that as $\mathbf{c_q}$ increases, then $\mathbf{A}$ increases, while $\mathbf{Q}$ decreases. This led to the formulation of the *query-cost interaction* hypothesis. While this is an intuitive prediction of search behaviour, the model makes a number of assumptions and as a result has a number of limitations.

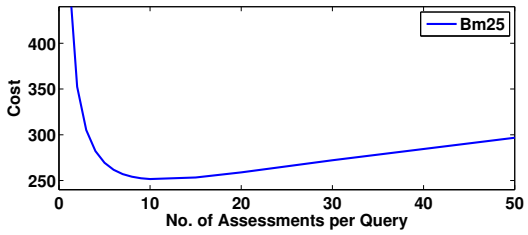

Figure 2: Plot of the Cost Function for the BM25.

3.1 Model Limitations

Three main assumptions of the approach were pointed out in [5]: (1) rationality and optimality, (2) the fixed interaction, and (3) the abstraction of the search interface. The first assumption is a common modelling assumption often employed (*c.f.* [9, 25, 30]). While it assumes that users will act in a rational way, it is widely acknowledged [2, 25, 30] that users are not perfect and are susceptible to biases (e.g. [42]). However, this does not mean that users do not try to act in such a manner. In [33, 38], they showed that users adapted their behaviour to the system to obtain the same level of gain. While in [27], they showed that users did tend to minimise effort when engaging with the scatter-gather system. These findings suggest that this is a reasonable assumption, but it is clear that there is scope to improve the model on this front. However, we leave this to further work and focus on addressing more pragmatic issues first.

Next there is the assumption that user interaction is fixed, i.e. that users will examine $\mathbf{A}$ documents per query. Clearly, users will not always examine the same number of documents for each query that they issue. In [2], it is acknowledged that user behaviour is not static nor is it fixed, and so they argued that this value represents the average number of documents examined per query. This approximation is similar to those made by numerous evaluation measures (i.e. p@k, NDCG@100, etc [32]), where the depth is fixed. However, it would be possible to generalise the model and use a distribution to represent the number of documents per query rather than the mean. Of course, this would significantly increase the complexity of the model which may not be necessary. In its current form, though, it is possible to overcome this limitation by comparing how the search behaviour, characterised by $\mathbf{A}^\star$ and $\mathbf{Q}^\star$, would change when moving from one situations to another. For example, when the cost of querying is low on one interface, but high on another, how would search behaviour change? This is often referred to as *comparative statics* in Economics and in the next section we use this method to generate specific hypotheses regarding search behaviour.

Finally, this model of the search interface is quite simplistic and reminiscent of the TREC abstraction of a search interface [34]. For instance, viewing the search result pages or inspecting snippets are not considered by the model. However, such interactions come at a cost to the user, and also helps them to decide what action to perform next. In this respect, the model currently assumes that these aspects do not play a significant role in shaping the user's behaviour. A similar problematic assumption has made by most evaluation measures as they ignore the influence of the interface [32, 37, 38]. The cost function also ignores these additional costs which are likely to have an impact on the usefulness and accuracy of the economic model. We seek to address these limitations of the model by using a more accurate model of the search interface.

3.2 Empirical Research

In addition to these assumptions, the study conducted in [5] showed mixed empirical evidence to support the theory. Specifically, they tested the *query-cost interaction* hypotheses using a between-subjects experimental design. On the TREC Aquaint collection with three Robust 2005 ad-hoc topic tasks, participants used a BM25-based search system but with different search interfaces. The three interfaces were **Condition 1**: a structured interface which slowed query entry (high query cost), **Condition 2**: a standard search box (medium query cost), and **Condition 3**: a standard search box along with eight query suggestions (low query cost). It was found that participants on the high cost interface issued fewer queries ($\mathbf{Q_1} = 19.4$) and examined more documents per query ($\mathbf{A_1} = 4.7$) than the other two conditions, thus supporting the theory. However, participants on the low cost interface, contrary to expectation, issued slightly fewer queries ($\mathbf{Q_3} = 31.2$) and examined more documents per query ($\mathbf{A_3} = 2.5$) than the medium cost interface (where $\mathbf{Q_2} = 35.0$ and $\mathbf{A_2} = 1.6$). It was hypothesised that the reason behind this finding was that participants on the low cost interface spent more time on the search results page considering the query suggestions; something that was not part of the initial model, i.e. time on search page. Also participants on the low cost interface had access to good quality query suggestions and so experienced higher levels of precision, on average.

These findings strongly motivate a revision of the theory, and question whether it can provide credible explanations of search behaviours. In this work, we shall develop a new model of search which is more realistic to determine whether

this can improve the ability of the model to describe, predict and explain search behaviours.

4. NEW ECONOMIC MODEL OF SEARCH

Given the aforementioned limitations, we shall revise the current model by modifying the gain and cost functions to be more intuitive and reflective of the actual search process. To make such refinements, we shall assume that the search interface is much like a standard web search interface consisting of a query box (or query area) and search button. When a query is issued to the IR system the search result page shows: (i) the number of search results, (ii) the current page number, (iii) a list of n results (usually $n = 10$ results per page) and (iv) a next and previous button. Each search result has a title (often shown as a blue link), a snippet from the document, along with the URL/domain. This style of interface is usually referred to as the ten blue links.

On this interface, the user can perform a number of interactions: (i) (re)query, (ii) examine the search results page, (iii) inspect individual result snippets, (iv) assess documents and (v) visit subsequent results pages. Each of these actions have an associated cost and so are likely to affect search behaviour. This model of the search interface is much like those assumed in [31, 34], where the evaluation measures developed in those works include the processing time and effort associated with viewing and examining snippets. For example, with Time Biased Gain (TBG) [34], the interactions the user performs are dependant on the perceived relevance of a snippet, the reading speed of users, and the probabilities of viewing/judging documents (and similarly with the U-measure [31]).

During the course of interaction, a user will pose a number of queries ($\mathbf{Q}$), examine a number of search result pages per query ($\mathbf{V}$), inspect a number of snippets per query ($\mathbf{S}$) and with some probability $\mathbf{p_a}$ assess a number of documents per query ($\mathbf{A}$). Each interaction has an associated cost where $\mathbf{c_q}$ is the cost of a query, $\mathbf{c_v}$ is the cost of viewing a page, $\mathbf{c_s}$ is the cost of inspecting a snippet, and $\mathbf{c_a}$ is the cost of assessing a document. Note that the costs could be time as in [34] or effort as in [31]. With this model of the search interface we can construct a new cost function that includes these additional variables and costs, such that the total cost of interaction is:

$$c(\mathbf{Q}, \mathbf{V}, \mathbf{S}, \mathbf{A}) = \mathbf{c_q}.\mathbf{Q} + \mathbf{c_v}.\mathbf{V}.\mathbf{Q} + \mathbf{c_s}.\mathbf{S}.\mathbf{Q} + \mathbf{c_a}.\mathbf{A}.\mathbf{Q} \quad (3)$$

This new cost function provides a richer representation of the costs incurred during the course of interaction. However, with the introduction of these additional variables it significantly increases the complexity of the model. In order to simplify the cost function, we will make a number of assumptions. First, we assume that $\mathbf{V}$ represents the average number of search pages viewed per query much like the assumption regarding the number of documents assessed, but somewhat weaker. This is because most of the time only one page of results is viewed. Consequently we will treat $\mathbf{V}$ as a constant $\mathbf{v}$, which represents the mean number of pages examined per query. However, it would be possible to encode the number of page views per query more precisely by using a step function based on the number of snippets viewed, representing the fixed cost incurred to load and view each page of results[2]. However, we leave this extension for further work. Second, we shall assume that the number of documents assessed will be proportional to the number of snippets viewed, and that users will inspect a snippet before examining a document, thus $\mathbf{S} \geq \mathbf{A}$. If we let the probability of assessing a document given the snippet be $\mathbf{p_a}$, then the expected number of assessments viewed per query would be $\mathbf{A} = \mathbf{S}.\mathbf{p_a}$. Substituting these values into the cost model, we obtained:

$$\mathbf{c}(\mathbf{Q}, \mathbf{V}, \mathbf{S}, \mathbf{A}) = \mathbf{c_q}.\mathbf{Q} + \mathbf{c_v}.\mathbf{v}.\mathbf{Q} + \mathbf{c_s}.\frac{\mathbf{A}}{\mathbf{p_a}}.\mathbf{Q} + \mathbf{c_a}.\mathbf{A}.\mathbf{Q} \quad (4)$$

We can now reduce the cost function to be dependent only on $\mathbf{A}$ and $\mathbf{Q}$, such that:

$$\mathbf{c}(\mathbf{Q}, \mathbf{A}) = (\mathbf{c_q} + \mathbf{c_v}.\mathbf{v}).\mathbf{Q} + \Big(\frac{\mathbf{c_s}}{\mathbf{p_a}} + \mathbf{c_a}\Big).\mathbf{A}.\mathbf{Q} \quad (5)$$

As we shall see later, this leads to some interesting insights into search behaviour given changes in cost. Now to make the model more intuitive, we propose a small change to the existing gain function, shown in Eq. 1, such that:

$$\mathbf{g}(\mathbf{Q}, \mathbf{A}) = \mathbf{k}.\mathbf{Q}^{\boldsymbol{\alpha}}.\mathbf{A}^{\boldsymbol{\beta}} \quad (6)$$

In the original model $\boldsymbol{\alpha}$ represented the relative efficiency of querying and assessing. Here we have decoupled this relationship and address an obvious problem: if the user issued $\mathbf{m}$ queries and examined $\mathbf{n}$ documents per query and all $\mathbf{n}$ documents per query were relevant, the gain would be $\mathbf{m} \times \mathbf{n}$. With the original gain function shown in Equation 1, it would not be possible to set $\boldsymbol{\alpha}$ such that $\mathbf{g}(\mathbf{Q}, \mathbf{A}) = \mathbf{m}.\mathbf{n}$. However, this revised function can cater for this situation by setting $\boldsymbol{\alpha} = \boldsymbol{\beta} = 1$. Furthermore, with this functional form it is possible to directly estimate the gain given a particular result list (or set of). Now that we have revised the model, we can now consider what this model says about search behaviour.

4.1 Optimal Search Behaviour

Using this model it is possible to determine what the optimal search behaviour (in terms of $\mathbf{Q}$ and $\mathbf{A}$) would be given the parameters of our model. To do this we assume that the objective of the user is to minimise the cost for a given level of gain (or alternatively, maximise their gain for a given cost). This occurs when the marginal gain equals the marginal cost. We can solve this optimisation problem with the following objective function (using a Lagrangian Multiplier λ):

$$\Delta = (\mathbf{c_q} + \mathbf{c_v}.\mathbf{v}).\mathbf{Q} + \Big(\frac{\mathbf{c_s}}{\mathbf{p_a}} + \mathbf{c_a}\Big).\mathbf{A}.\mathbf{Q} - \lambda\Big(\mathbf{k}.\mathbf{Q}^{\boldsymbol{\alpha}}.\mathbf{A}^{\boldsymbol{\beta}} - \mathbf{g}\Big)$$

where the goal is to minimise the cost subject to the constraint that the amount of gain is $\mathbf{g}$. This is the analytical analogy to the graphical solution presented in subsection 3. By taking the partial derivatives, we obtain:

$$\frac{\partial \Delta}{\partial \mathbf{A}} = \big(\frac{\mathbf{c_s}}{\mathbf{p_a}} + \mathbf{c_a}\big).\mathbf{Q} - \lambda.\mathbf{k}.\boldsymbol{\beta}.\mathbf{Q}^{\boldsymbol{\alpha}}.\mathbf{A}^{\boldsymbol{\beta-1}} \quad (7)$$

and:

$$\frac{\partial \Delta}{\partial \mathbf{Q}} = \mathbf{c_q} + \mathbf{c_v}.\mathbf{v} + \big(\frac{\mathbf{c_s}}{\mathbf{p_a}} + \mathbf{c_a}\big).\mathbf{A} - \lambda.\mathbf{k}.\boldsymbol{\alpha}.\mathbf{Q}^{\boldsymbol{\alpha-1}}.\mathbf{A}^{\boldsymbol{\beta}} \quad (8)$$

[2]The step function would be such that, the number of pages viewed $\mathbf{V}$ would be equal to the number of snippets viewed divided by the number of snippets shown per page ($\mathbf{n}$) rounded up to the nearest integer, i.e. $\lceil \frac{\mathbf{S}}{\mathbf{n}} \rceil$.

Setting these both to zero, and then solving, we obtain the following expressions for the optimal number of assessments per query $\mathbf{A}^\star$:

$$\mathbf{A}^\star = \frac{\boldsymbol{\beta}.(\mathbf{c_q} + \mathbf{c_v}.\mathbf{v})}{(\boldsymbol{\alpha} - \boldsymbol{\beta}).\left(\frac{\mathbf{c_s}}{\mathbf{p_a}} + \mathbf{c_a}\right)} \tag{9}$$

and the optimal number of queries $\mathbf{Q}^\star$:

$$\mathbf{Q}^\star = \sqrt[\boldsymbol{\alpha}]{\frac{\mathbf{g}}{\mathbf{k}.\mathbf{A}^{\boldsymbol{\beta}}}} \tag{10}$$

Using this analytical solution we can now generate a number of testable hypotheses about search behaviour by considering how interaction will change when specific parameters in the model increase or decrease. It should be noted that we do not believe that users will act optimally. However, like previous work [2, 9, 25], we do believe that users will try to act in such a manner. We therefore expect that actual search behaviour will be in line with the predicted search behaviour. With that in mind, we present eight hypotheses regarding search behaviour of which three relate to how performance affects behaviour, four relate to how the costs affect behaviour and the remaining hypothesis is based on the probability of assessing documents. These hypotheses are made assuming that the user wishes to obtain a fixed level of gain and that all other things are equal.

4.2 Performance and Interaction

In this subsection we define the three performance-based hypotheses. The **k-performance-interaction hypothesis** is as follows: as $\mathbf{k}$ increases the number of queries issued will decrease, while the number of documents examined per query will remain constant for a fixed level of gain (as shown in Figure 3). From the equations above we can see that $\mathbf{A}^\star$ is independent of $\mathbf{k}$, which is why $\mathbf{A}$ remains constant. While if we examine Equation 10, we can see that as $\mathbf{k}$ tends to infinity, then $\mathbf{Q}^\star$ tends to zero. Of course, in practice at least one query needs to be submitted to the system (i.e. $\mathbf{Q} \geq 1$). In [2], $\mathbf{k}$ is said to represent the efficiency of the user or the system in identifying or returning relevant documents. So $\mathbf{k}$ could represent: (i) the probability of issuing a query that returns relevant documents, (ii) the probability of a user selecting a relevant document from the ranking, (iii) the precision of the result list, or (iv) some combination of these factors. This is an open question.

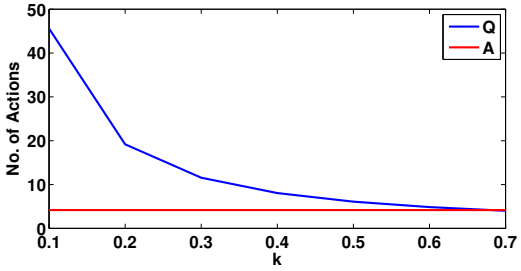

Figure 3: Plot of the $\mathbf{A}^\star$ and $\mathbf{Q}^\star$ as k changes.

The **$\boldsymbol{\alpha}$-performance-interaction hypothesis** can be formulated as follows: as $\boldsymbol{\alpha}$ increases, the number of assessments per query will decrease, while the the number of queries will increase. The **$\boldsymbol{\beta}$-performance-interaction hypothesis** is: as $\boldsymbol{\beta}$ increases, the number of assessments per query will increase, while the number of queries will decrease (as shown in Figure 4). Here $\boldsymbol{\beta}$ represents the quality of the ranked list, while $\boldsymbol{\alpha}$ represents the quality of the queries issued. There is an interesting relationship between these parameters. Since $\mathbf{A}^\star$ must be positive then the fraction $\frac{\boldsymbol{\beta}}{\boldsymbol{\alpha}-\boldsymbol{\beta}}$ also needs to be positive (assuming the costs $\mathbf{c_q}$ and $\mathbf{c_a}$ are also positive). This implies that $\boldsymbol{\alpha}$ needs to be greater than $\boldsymbol{\beta}$. Furthermore, as $\boldsymbol{\beta}$ tends to $\boldsymbol{\alpha}$, then the fraction $\frac{\boldsymbol{\beta}}{\boldsymbol{\alpha}-\boldsymbol{\beta}}$ tends to infinity, suggesting that assessing is preferable to querying. Restated, it is preferable to continue assessing the current ranked list in order to find more relevant documents, rather than issuing another query. For example, if a user posed the perfect query which returned all the relevant documents, then subsequent queries would be redundant and only increase the overall costs. Another interesting observation is that if $\boldsymbol{\alpha}$ equals one or tends to one, then the queries issued are likely to be independent of each other. That is, they would be returning different relevant documents (or sets of), whereas if $\boldsymbol{\alpha}$ equals zero or tends to zero, then the queries issued are likely to be similar and so would be returning relevant documents that have already been observed (and thus not contributing to the overall gain).

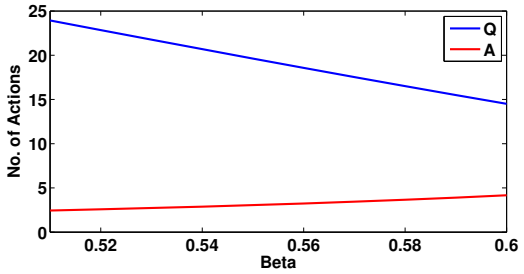

Figure 4: Plot of $\mathbf{A}^\star$ and $\mathbf{Q}^\star$ as $\boldsymbol{\beta}$ changes.

4.3 Cost and Interaction

Next, we define the series of cost-based hypotheses regarding search behaviour. The **Query-cost-interaction hypothesis**, which has already been asserted in [5], is as follows: as the cost of a query $\mathbf{c_q}$ increases, the number of documents assessed per query will increase, while the number of queries issued will decrease (as shown in Figure 5). It should be clear from the Equation 9 that this is the case because as $\mathbf{c_q}$ approaches infinity, $\mathbf{A}^\star$ also approaches infinity. In turn, the number of queries issued will decrease, because as $\mathbf{A}$ becomes larger, $\mathbf{Q}^\star$ will tend to zero. As previously mentioned, $\mathbf{Q}$ must be equal to one or greater.

Similarly, we can formulate the **page-cost-interaction hypothesis**: as the cost of viewing a page increases, the number of documents assessed per query will increase, while the number of queries issued will decrease.

The **Assessment-cost-interaction hypothesis** is: as the cost of an assessment increases, the number of documents assessed per query will decrease, while the number of queries issued will increase. Since the assessment cost $\mathbf{c_a}$ appears in the denominator in Equation 9 then any increase will reduce the number of assessments.

Similarly for the **Snippet-cost-interaction hypothesis** is: as the cost of processing snippets increases, the number of documents assessed per query will decrease, while the number of queries issued will increase.

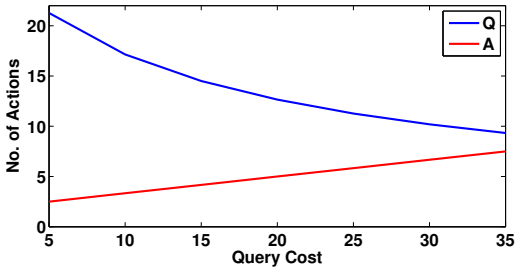

Figure 5: Plot of $\mathbf{A}^\star$ and $\mathbf{Q}^\star$ as query cost changes.

Finally, the **Assessment-probability-interaction hypothesis** can be stated as follows: as the probability of assessing increases given the result snippet, the number of documents assessed increases, while the number of queries issued decreases (as shown in Figure 6). Here, if a user examines every document in the ranked list, then $\mathbf{p_a}$ would equal one meaning that they also examine every snippet.

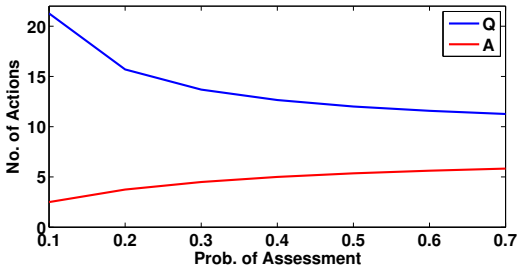

Figure 6: Plot of $\mathbf{A}^\star$ and $\mathbf{Q}^\star$ as the probability of assessment changes.

4.4 Reconsidering Past Results

Now lets consider whether this new model can provide an explanation for the search behaviour observed in [5] and previously described in subsection 3.2. Recall that in their experiment, the page costs in the suggestion interface were higher than on the standard interface (while the query costs were similar). It was also the case that when compared to the standard interface, fewer queries were issued, and more documents per query were examined (which was inconsistent with their expectations). However, given the *page-cost-interaction* hypothesis this observation is to be expected as participants on the suggestion interface spent longer per page. Alternatively, the change in behaviour may have been due to differences in performance as participants on the suggestion interface experienced higher levels of precision (and thus an increase in $\boldsymbol{\beta}$). According to the $\boldsymbol{\beta}$-*interaction* hypothesis, such a change would also result in a similar observation. Consequently, the refined model provides two credible and possible explanations of the observed search behaviour, which the previous model could not.

5. METHOD

In this section, we will undertake an analysis of the search behaviours of users to see whether actual search behaviours are consistent with the hypotheses generated in the previous section. To perform this analysis, we used the search logs from the study conducted by Azzopardi *et. al.* [5], which was previously described in Subsection 3.2. The search logs contained the transactions for 36 participants on the three different conditions: (**1**) structured search interface, (**2**) standard interface and (**3**) suggestion interface, where there were 12 participants per condition and each participant performed three search tasks from the TREC Aquaint Collection (344: Abuses of E-mail, 347: Wildlife Extinction, and 435: Curbing Population Growth).

For this analysis, we used time (in seconds) to represent the cost of the various interactions as done in [5]. We also focused on analysing the search behaviour on a topic by topic basis (which was not done in [5]). There are two main reasons for this: (i) topics are a major source of variation where users are likely to face different difficulties depending on the topic (experiencing different costs and different levels of performance because of the topic), and (ii) the model is session based representing the interaction when searching for a specific topic. Furthermore, we performed the analysis on each condition, though aggregating all conditions together resulted in similar findings.

From the search logs it was possible to extract for each user on a given topic: the number of queries issued ($\mathbf{Q}$), the number of pages examined ($\mathbf{V}$), the number of snippets viewed ($\mathbf{S}$) and the number of documents examined ($\mathbf{A}$). It was also possible to extract the time spent issuing a query ($\mathbf{c_q}$) and the time spent examining a document ($\mathbf{c_d}$). The time per snippet and time per page was not possible to directly extract from the log data. However, we could extract the time spent on the search engine result page ($\mathbf{t_{serp}}$) from which it was possible to estimate $\mathbf{c_s}$ and $\mathbf{c_v}$. This was performed by assuming that $\mathbf{t_{serp}} = \mathbf{c_s}.\mathbf{S} + \mathbf{c_v}.\mathbf{V}$, where $\mathbf{V}$ was set to the average number of pages (and thus a constant) and we fitted the data to this function. Table 1 reports the estimated values and the goodness of fit $\mathbf{r}$.

The next set of parameters that we estimated were $\mathbf{k}$ and $\boldsymbol{\beta}$. This was performed on a query-by-query basis. For each query, we calculated the Cumulative Gain [19] for each rank from 1 to 50. We then estimated the parameters for $\mathbf{k}$ and $\boldsymbol{\beta}$ using a least squared approximation. Note that if the cumulative gain across all ranks was zero for a given query then both $\mathbf{k}$ and $\boldsymbol{\beta}$ were zero. Table 1 shows the estimates for $\mathbf{k}$ and $\boldsymbol{\beta}$ for each condition and topic given the condition. Further note that when reporting significance we used a one-way ANOVA test for comparing means.

6. ANALYSIS

Now we shall examine on each of the conditions, whether the change in querying and assessing is consistent with the predicted trends. For the this part of the analysis, we have plotted the change in querying and assessing for the different model parameters. Note that we are averaging over sessions/topics for each condition - so we are essentially assuming that all the other variables are equal. While this is not the case in practice, we hope that the predicted trend is strong enough to be observed.

6.1 Performance and Interaction

$\mathbf{k}$-interaction hypotheses: Figure 7 shows the plots for each condition for each topic when $\mathbf{k}$ (top) and $\boldsymbol{\beta}$ (bottom) changes, and how $\mathbf{A}$ and $\mathbf{Q}$ respond. If we first consider the $\mathbf{k}$-plots, then we can see for conditions 2 (mid) and 3 (right), that as $\mathbf{k}$ increases, $\mathbf{Q}$ decreases – which was implied by the $\mathbf{k}$-interaction hypothesis. However, the number of documents assessed also tends to increase – which was not implied by the hypothesis. Furthermore, if we examine the

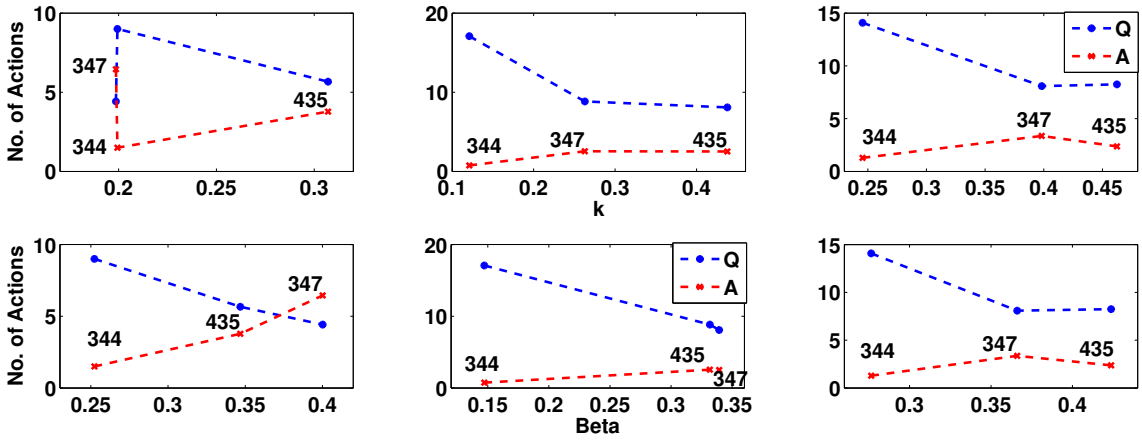

Figure 7: Top: Plots of A and Q as k changes. Bottom: Plots of A and Q as β changes. Left: Condition 1, Mid: Condition 2, Right: Condition 3.

Topic	Cond.	Q	V	S	D	p_a	c_q	c_d	c_s	c_v	r	k	b
All	1	19.1	1.69	12.3	3.32	0.179	26.5	14.6	2.56	19.0	0.557	0.231	0.315
	2	34.0	0.92	7.02	1.64	0.145	15.2	16.9	2.68	7.8	0.494	0.233	0.241
	3	30.4	1.98	10.1	2.13	0.153	13.8	16.5	1.68	14.4	0.614	0.345	0.340
344	1	9.0	1.35	8.51	1.5	0.119	24.3	16.9	2.73	17.0	0.551	0.199	0.252
	2	17.1	0.82	5.38	0.75	0.083	14.4	16.8	2.34	9.8	0.323	0.121	0.147
	3	14.1	1.58	8.18	1.28	0.101	14.6	16.8	1.44	15.1	0.535	0.246	0.276
347	1	4.4	2.09	19.0	6.45	0.260	30.3	11.2	2.09	23.7	0.557	0.199	0.400
	2	8.8	1.12	9.09	2.55	0.194	15.4	14.9	3.01	4.2	0.790	0.263	0.332
	3	8.1	2.69	12.4	3.36	0.212	12.7	15.0	1.83	12.6	0.819	0.398	0.366
435	1	5.7	1.77	13.0	3.78	0.213	27.0	15.0	2.65	18.3	0.567	0.307	0.347
	2	8.1	0.87	8.25	2.52	0.223	16.6	19.3	3.02	7.4	0.533	0.438	0.339
	3	8.3	1.77	10.9	2.37	0.184	13.6	17.9	1.92	14.8	0.546	0.462	0.424

Table 1: A summary of the model parameters, where the mean is reported for each parameter.

plot for condition 1 (top, left plot), this trend does not appear to hold. For topics 344 and 347, the **k** values are equal, yet the **A** and **Q** values are quite different. This may be because not all other things are equal. If we look at the β-plot for condition 1 (bottom, left plot), the β values for these two topics are quite different: $\beta_{344} = 0.25$ and $\beta_{347} = 0.4$. This may explain the difference observed on the **k**-plot. The β-interaction hypothesis states that an increase in β, would result in more documents being examined (i.e. **A** $\uparrow$) and less queries being issued (i.e. **Q** $\downarrow$). When we compare the two topics on condition 1, we see that for topic 344, **A** is 1.5 and **Q** is 9, while for 347, **A** increases to 6.45 and **Q** decreases to 4.42; which falls in line with the β-interaction hypothesis.

β-interaction hypotheses: In terms of the β-plots, we see that across each condition the trend lines follow the β-interaction hypothesis: as β increases, we see that **A** increases, while **Q** tends to decrease. This is despite the fact that we are averaging over all sessions/users, where many other factors are varying. Intuitively, this suggests that the performance of the query characterised by β plays a major role in shaping the interaction of the user and that β appears to dominate over **k**, suggesting that there is an ordering to the importance of these hypotheses.

In Table 2, we have performed a deeper analysis with respect to β to determine if the number of documents examined was affected by β. For each condition, we grouped the queries issued into those where β was equal to zero (i.e. no TREC relevant results were returned) and another group where β was greater than zero (i.e. TREC relevant results were returned). For each group, we computed the number of documents examined and used an ANOVA test to determine if the means were different. For most (all but one) conditions/topics the difference was significantly different. This further shows that β plays a significant role in shaping interaction.

Note we have excluded reporting on the α-interaction hypothesis as there were numerous ways in which α could have been estimated. However, each was open to interpretation, and so a subject of investigation in its own right. Thus, we leave this for further work.

6.2 Cost and Interaction

Query-cost-interaction hypothesis: From the top plots in Figure 8, we can see that for conditions 1 and 2, as the cost of a query increases, the number of assessments increases, while the number of queries decreases. This is in line with the hypothesis. However, in condition 3, the trend was the opposite. It may be that for this condition the performance factors are overriding or hiding the influence of the query cost. For topic 344, the **k** and β parameters are very low compared to the other topics for this condition. Furthermore, the cost of a query between topics on this condition were not significantly different ($F(2,33)=0.97$, $p=0.38$), suggesting that other factors could be responsible for the observed trend. Although not shown due to space

Topic	Cond.	$\mathbf{A}(\boldsymbol{\beta} = \mathbf{0})$	$\mathbf{A}(\boldsymbol{\beta} > \mathbf{0})$	$\mathbf{p}(\mathbf{b} > \mathbf{0})$	Q
All	**1**†	1.50±3.4	5.4±6.9	0.47	229
	2†	0.64±1.5	3.0±4.1	0.42	408
	3†	0.86±1.9	3.2±4.8	0.55	365
344	**1**†	0.65±1.1	2.8±3.6	0.39	108
	2†	0.51±1.2	1.5±2.2	0.26	205
	3†	0.49±1.1	2.3±4.2	0.44	169
347	**1**†	2.00±2.7	10.0±9.9	0.55	53
	2†	0.96±2.3	3.8±5.1	0.56	106
	3⋆	1.70±3.2	4.2±6.3	0.66	97
435	**1**	2.80±5.8	4.6±5.2	0.54	68
	2†	0.77±1.4	3.7±3.9	0.60	97
	3†	1.00±1.7	3.1±3.3	0.65	99

Table 2: The mean and standard deviation of the number of documents assessed per query for each topic and all topics for each condition, along with the probability of issuing a query where $\mathbf{b} > \mathbf{0}$ and the number of queries Q. ⋆ (†) indicates a significance difference at $\mathbf{p} < 0.05$ ($\mathbf{p} < 0.01$).

constraints, a similar trend was also observed for the **page-cost-interaction hypothesis**.

Assessment-cost-interaction hypothesis: The bottom plots in Figure 8 show the trends given the document cost. For condition 1, the expected behaviour was observed, while for the other conditions the trend is rather mottled. This suggests that the document cost interaction hypothesis may not hold. To inspect this further, we compared the document costs between topics and found that there were no significant differences in any of the conditions (1: F(2,33)=1.88, p=0.16, 2: F(2,33)=2.07, p=0.13, and 3: F(2,33)=0.88, p=0.42). This suggests that the observed differences were due to other factors and not document cost (see below).

Assessment-Probability-Interaction hypothesis: The top plots in Figure 9 show how the interaction changed as the probability of assessment increased. The trend across all conditions was similar, and consistent with the hypothesis that as the probability of assessment increased, the number of documents assessed per query increased, while the number of queries decreased. We also found although not shown due to space constraints, that the data was also consistent with the **snippet-cost-interaction** hypothesis.

Now, we turn our attention back to the document costs, and how the trend did not meet our initial expectations. In Figure 9, the bottom plots show the change in **A** and **Q** for the expression $\frac{\mathbf{c_s}}{\mathbf{p_a}} + \mathbf{c_a}$. Since the $\mathbf{p_a}$ tends to be small and is less than one, the fraction $\frac{\mathbf{c_s}}{\mathbf{p_a}}$ is likely to be a greater influence than $\mathbf{c_a}$. Consequently, as $\frac{\mathbf{c_s}}{\mathbf{p_a}} + \mathbf{c_a}$ increases, we would expect to see the number of documents assessed decrease, while the number of queries would increase. For each condition, we see that this trend is observed. This suggests that the probability of assessment played a larger role in determining search behaviour for the range of values of $\mathbf{c_a}$ and $\mathbf{c_s}$ in this experimental data. This is also confirmed by follow-up significance tests which shows that the mean $\mathbf{p_a}$ within each condition was significantly different (1: F(2,33)=13.1, p<0.001, 2: F(2,33)=26.7, p<0.001, and 3: F(2,33)=13.6, p<0.001). However, if $\mathbf{c_a}$ was substantially larger than $\frac{\mathbf{c_s}}{\mathbf{p_a}}$, the document cost would dominate the expression and we would then expect $\mathbf{c_a}$ to have a greater influence on search behaviour. An interesting direction for future work would be to examine whether this is the case or not when the cost of assessing a document is varied significantly between different conditions.

7. DISCUSSION AND CONCLUSION

In this paper, we have developed a new economic model of the search process. This new model has led to a number of insights regarding search behaviour, which have been expressed as a series of eight interaction hypotheses. The model provides a better account of past empirical observations, and when we analysed the user interactions of thirty-six users we found trends consistent with the theory. In the cases where the trends bucked against our expectations, it was revealed that certain variables had a greater influence on search behaviour, thus explaining these deviations. From our analysis, it appears that the performance variable $\boldsymbol{\beta}$ plays the most important part in determining the search behaviour, and then the probability of assessing, $\mathbf{p_a}$, and the cost of inspecting a snippet, $\mathbf{c_s}$. However, the ordering depends on the magnitude of the costs and probabilities. Furthermore, this exposition shows how the revised theory can provide credible explanations of observed behaviours; providing a compact representation of expected search behaviour, which practitioners and researchers can draw upon when designing experiments and in explaining and understanding observed search behaviour.

While this is a promising step forward, the model is far from perfect, and there are many ways in which it could be improved. Further work is needed to make the model even more realistic of the search process, removing approximations and including scope to include constraints and/or known biases [42]. The model also does not include the probability of issuing a good query or the probability of correctly marking a document relevant, either. Such probabilities are included in the users models of TBG [34] and U-Measure [31], for instance. So it would be a natural step to include these within the economic model, or look to integrate these measures into the model. Also, while we have generated a number of testable hypotheses, more work is required to empirically test and examine whether these hypotheses hold in practice on and across more topics, on different types of tasks, and under various conditions. However, now that we have developed these hypotheses it has paved the way forward for future experimentation that is theoretically underpinned - something that was found lacking in many past works [22]. Another future direction is to develop IFT [25] and the iPRP [16] in order to generate a similar set of hypotheses about search behaviour in order to compare and contrast the different theoretical frameworks.

Acknowledgments I would like to thank Keith van Rijsbergen, Guido Zuccon and David Elswelier for their helpful and insightful comments and discussions. Also, I would like to thank the participants at ADCS 2013 for their feedback on [3] and the reviewers of this paper for their supportive comments and suggestions. All of which has helped to improve this work and provide many directions for future work.

8. REFERENCES

[1] C. W. Axelrod. The economic evaluation of information storage and retrieval systems. *Information Processing & Management*, 13(2):117–124, 1977.

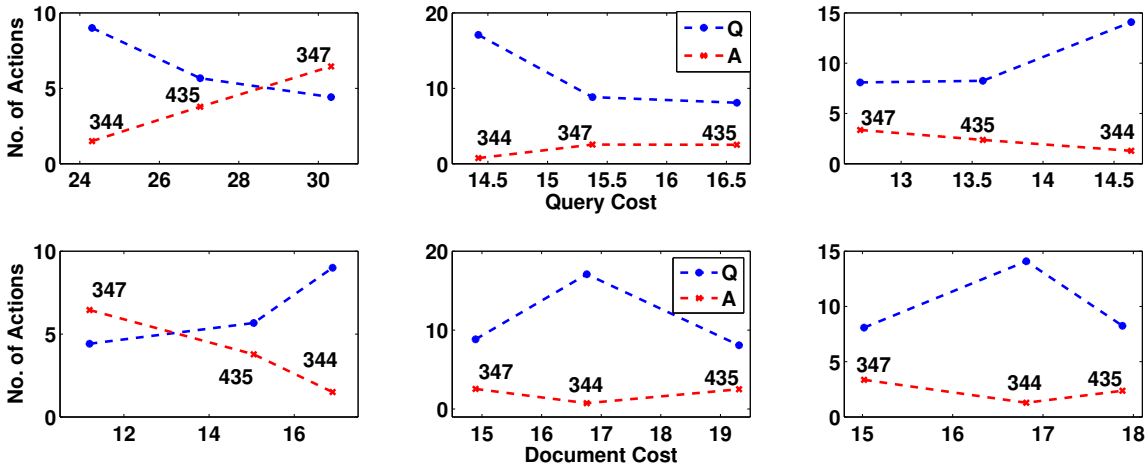

Figure 8: Top: Plots of A and Q as the cost of a query changes. Bottom: Plots of A and Q as the cost of a assessing a document changes. Left: Condition 1, Mid: Condition 2, Right: Condition 3.

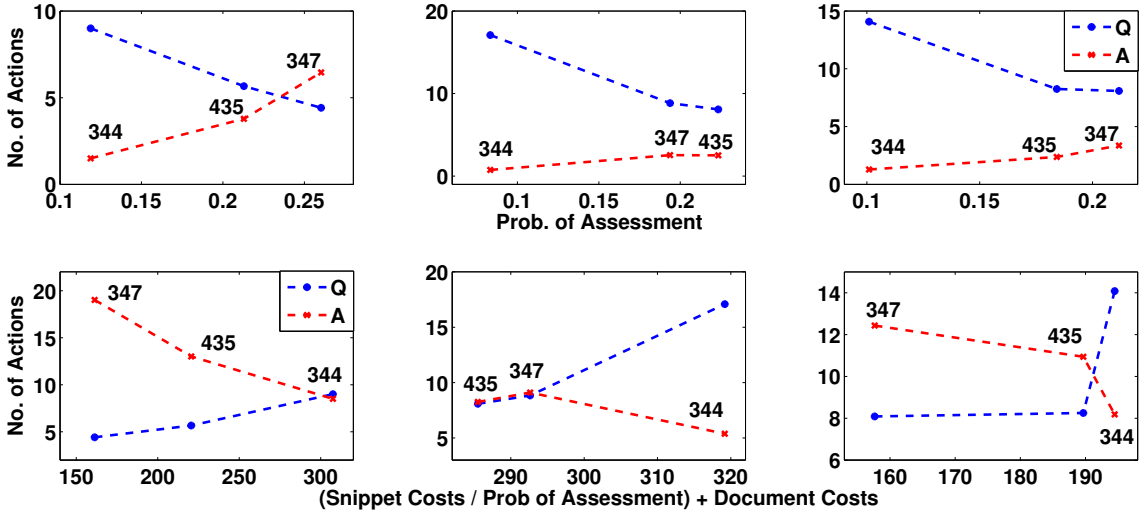

Figure 9: Top: Plots of the A and Q as the probability of assessment changes. Bottom: Plots of the A and Q as the combined value changes. Left: Condition 1, Mid: Condition 2, Right: Condition 3.

[2] L. Azzopardi. The economics in interactive information retrieval. In *Proceedings of the 34th ACM conference on research and development in information retrieval (SIGIR)*, pages 15–24, 2011.

[3] L. Azzopardi. Economic models of search. In *Proceedings of the 18th Australasian Document Computing Symposium*, ADCS '13, pages 1–1, 2013.

[4] L. Azzopardi, K. Järvelin, J. Kamps, and M. D. Smucker. Sigir 2010 report workshop on the simulation of interaction. *SIGIR Forum*, 44:35–47, 2011.

[5] L. Azzopardi, D. Kelly, and K. Brennan. How query cost affects search behavior. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in IR*, pages 23–32, 2013.

[6] F. Baskaya, H. Keskustalo, and K. Järvelin. Time drives interaction: simulating sessions in diverse searching environments. In *Proceedings of the 35th ACM conference on research and development in information retrieval (SIGIR)*, pages 105–114, 2012.

[7] F. Baskaya, H. Keskustalo, and K. Järvelin. Modeling behavioral factors ininteractive information retrieval. In *Proceedings of the 22nd ACM international conference on Conference on information and knowledge management*, pages 2297–2302, 2013.

[8] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.

[9] M. J. Bates. Training and education for online. chapter Information search tactics, pages 96–105. Taylor Graham Publishing, London, UK, UK, 1989.

[10] N. J. Belkin. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42:47–54, 2008.

[11] N. J. Belkin, R. N. Oddy, and H. M. Brooks. Ask for information retrieval: part i: background and theory; part ii: results of a design study. *Journal of Documentation*, 38(2) 61-71 and 38(3) 145-164, 1982.

[12] U. Birchler and M. Butler. *Information Economics*. Routledge, 1st edition edition, 2007.

[13] J. Brutlag. Speed matters for google web search. In *Technical Report,2009, Retrieved online at http://googleresearch.blogspot.com/2009/06/speed-matters.html*.

[14] M. D. Cooper. A cost model for evaluating information retrieval systems. *Journal of the American Society for Information Science*, pages 306–312, 1972.

[15] S. Erdelez. Information encountering: a conceptual framework for accidental information discovery. In *Proceedings of an international conference on Information seeking in context*, pages 412–421, 1997.

[16] N. Fuhr. A probability ranking principle for interactive information retrieval. *Information Retrieval*, 11(3):251–265, 2008.

[17] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer-Verlag New York, Inc., 2005.

[18] K. Järvelin. Ir research: systems, interaction, evaluation and theories. *ACM SIGIR Forum*, 45(2):17–31, 2012.

[19] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20:2002, 2002.

[20] K. Järvelin and T. D. Wilson. On conceptual models for information seeking and retrieval research. *Information Research*, 9(1):9–1, 2003.

[21] D. Kelly, V. D. Dollu, and X. Fu. The loquacious user: a document-independent source of terms for query expansion. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 457–464, 2005.

[22] D. Kelly and C. Sugimoto. A systematic review of interactive information retrieval evaluation studies, 1967-2006. *Journal of the American Society for Information Science and Tech.*, 64(4):745–770, 2013.

[23] H. Keskustalo, K. Järvelin, A. Pirkola, T. Sharma, and M. Lykke. Test collection-based ir evaluation needs extension toward sessions — a case of extremely short queries. In *Proc. of the 5th AIRS*, pages 63–74, 2009.

[24] C. C. Kuhlthau. Developing a model of the library search process: Cognitive and affective aspects. *RQ*, pages 232–242, 1988.

[25] P. Pirolli and S. Card. Information foraging. *Psychological Review*, 106:643–675, 1999.

[26] P. Pirolli and W. T. Fu. Snif-act: a model of information foraging on the www. In *Proceedings of the 9th International Conference on User Modeling*, pages 45–54, 2003.

[27] P. Pirolli, P. Schank, M. Hearst, and C. Diehl. Scatter/gather browsing communicates the topic structure of a very large text collection. In *Proceedings of the ACM SIGCHI Conference*, pages 213–220, 1996.

[28] S. E. Robertson. The probability ranking principle in ir. *Journal of documentation*, 33(4):294–304, 1977.

[29] D. H. Rothenberg. An efficiency model and a performance function for an ir system. *Information Storage and Retrieval*, 5(3):109 – 122, 1969.

[30] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT/SIGCHI*, pages 269–276, 1993.

[31] T. Sakai and Z. Dou. Summaries, ranked retrieval and sessions: A unified framework for information access evaluation. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in IR*, pages 473–482, 2013.

[32] M. Sanderson. *Test collection based evaluation of information retrieval systems*. FNTIR, 2010.

[33] C. L. Smith and P. B. Kantor. User adaptation: good results from poor systems. In *Proceedings of the 31st ACM conference on research and development in information retrieval (SIGIR)*, pages 147–154, 2008.

[34] M. D. Smucker and C. L. Clarke. Time-based calibration of effectiveness measures. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in IR*, pages 95–104, 2012.

[35] G. J. Stigler. The economics of information. *The Journal of Political Economy*, 69(3):213–225, 1961.

[36] S. Sushmita, H. Joho, M. Lalmas, and R. Villa. Factors affecting click-through behavior in aggregated search interfaces. In *Proc. of the 19th ACM International Conference on Information and Knowledge Management*, pages 519–528, 2010.

[37] A. Turpin and W. Hersh. User interface effects in past batch versus user experiments. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '02, pages 431–432, 2002.

[38] A. H. Turpin and W. Hersh. Why batch and user evaluations do not give the same results. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in IR*, SIGIR '01, pages 225–231, 2001.

[39] H. R. Varian. *Intermediate microeconomics: A modern approach*. W.W. Norton, New York:, 1987.

[40] H. R. Varian. Economics and search. *SIGIR Forum*, 33(1):1–5, 1999.

[41] J. Wang and J. Zhu. Portfolio theory of information retrieval. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 115–122. ACM, 2009.

[42] R. White. Beliefs and biases in web search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3–12, 2013.

[43] R. W. White, S. T. Dumais, and J. Teevan. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, pages 132–141, 2009.

[44] R. W. White and D. Morris. Investigating the querying and browsing behavior of advanced search engine users. In *Proc. of the 30th ACM conference on research and development in IR*, pages 255–262, 2007.

[45] R. W. White, I. Ruthven, J. M. Jose, and C. J. V. Rijsbergen. Evaluating implicit feedback models using searcher simulations. *ACM Trans. Inf. Syst.*, 23(3):325–361, 2005.

[46] T. D. Wilson. Human information behavior. *Informing science*, 3(2):49–56, 2000.